

Variational Inequalities and Improved Convergence Rates for Tikhonov Regularisation on Banach Spaces

Markus Grasmair

Computational Science Center
University of Vienna
Nordbergstr. 15
A-1090 Vienna, Austria

July 13, 2011

Abstract

In this paper we derive higher order convergence rates in terms of the Bregman distance for Tikhonov like convex regularisation for linear operator equations on Banach spaces. The approach is based on the idea of variational inequalities, which are, however, not imposed on the original Tikhonov functional, but rather on a dual functional. Because of that, the approach is not limited to convergence rates of lower order, but yields the same range of rates that is well known for quadratic regularisation on Hilbert spaces.

1 Introduction

Classical results in the theory of quadratic Tikhonov regularisation on Hilbert spaces show that the asymptotic quality of the regularised solution of an ill-posed linear operator equation $Ax = y$ depends strongly on the smoothness of the true solution x^\dagger with respect to the operator $A: X \rightarrow Y$: If x^\dagger satisfies the range condition

$$x^\dagger \in \text{Ran}(A^* A)^\nu \quad (1)$$

for some $0 < \nu \leq 1$ and the accuracy of the right hand side of the equation is of order δ , then one obtains, with a suitable choice of the regularisation parameter, an accuracy of the regularised solution of order $\delta^{2\nu/(2\nu+1)}$. More precisely, denote by $y^\delta \in Y$ any right hand side satisfying $\|y - y^\delta\| \leq \delta$, and by $x_\alpha^\delta \in X$ with $\alpha > 0$ and $\delta > 0$ any element satisfying

$$x_\alpha^\delta = \arg \min \{ \|Ax - y^\delta\|^2 + \alpha \|x\|^2 : x \in X \}$$

for some y^δ . Then a choice of the regularisation parameter α of order $\alpha \sim \delta^{2/(2\nu+1)}$ implies that $\|x_\alpha^\delta - x^\dagger\| = O(\delta^{2\nu/(2\nu+1)})$ (see for instance [7] or [12]).

The situation becomes more complicated in the case where the regularisation term is not the squared Hilbert space norm and, even worse, the regularisation

takes place only in a Banach space and not a Hilbert space setting. Then Tikhonov regularisation consists in the minimisation of a functional

$$\mathcal{T}_\alpha(x; y^\delta) = \|Ax - y^\delta\|^2 + \alpha\mathcal{R}(x),$$

where \mathcal{R} is some convex and lower semi-continuous functional on X . The first results that treated convergence rates for non-quadratic regularisation remained close to the quadratic case. In [19] and [4], quadratic regularisation with convexity constraints was discussed, leading to conditions and results that are very similar to the unconstrained case. The first result that transcended the Hilbert space setting was the treatise [8], where convergence rates for maximum entropy regularisation were derived. The main argument, however, was a translation of the non-quadratic problem to an equivalent quadratic problem on a Hilbert space.

One of the main difficulty in the generalisation of the convergence rates result to general convex regularisation methods is that the norm on X , in which the convergence rate was usually measured, need not be related to the regularisation term that should imply the convergence. For this reason, it was argued in [3] that it is more natural to measure the quality of the approximation not in terms of the norm but rather in terms of the Bregman distance with respect to the regularisation term \mathcal{R} . Then it is easy to derive, for a parameter choice $\alpha \sim \delta$, a convergence rate of order δ provided the range condition $\text{Ran } A^* \cap \partial\mathcal{R}(x^\dagger) \neq \emptyset$ holds. Here $\partial\mathcal{R}(x^\dagger)$ denotes the sub-differential of the convex functional \mathcal{R} at x^\dagger . This result corresponds to the case $\nu = 1/2$ in (1). In addition, a result with the improved range condition $\text{Ran } A^* A \cap \partial\mathcal{R}(x^\dagger) \neq \emptyset$ with Y being a Hilbert space has been derived in [22] (see also [20, 21] for more general results).

The above results on convergence rates for regularisation on Banach spaces correspond to the cases $\nu = 1/2$ or $\nu = 1$ in (1); none of the intermediate cases was treated in [3] or [22]. There are two main reasons: First, an intermediate range condition is only possible if fractional powers of A can be defined, which basically limits the theory to Hilbert spaces, and, second, the classical proofs of the intermediate convergence rates rely on the possibility of writing x_α^δ as the solution of a linear operator equation.

An alternative approach for the derivation of convergence rates in a quadratic Hilbert space setting was introduced in [16] (see also [23], where a similar idea was used in a slightly different context) under the headings of *approximate source conditions* and *distance functions*. This approach was used with success in [14] (note also the preliminary results in [13]), leading to the whole range of estimates also available for the quadratic case. Another alternative, *variational inequalities*, arised in [17] (see also [24, Sec. 3.2]) from the search for a natural way of extending convergence rates results to non-linear problems. In [2, 11] it was shown that these variational inequalities can be used for covering the whole range of *slow* convergence rates corresponding to the cases $0 < \nu \leq 1/2$ in (1). The two concepts were also combined in [10] as *approximate variational inequalities*, but the obtained rates were not different from those in [2, 11].

In [9], the different new concepts for obtaining (slow) convergence rates for convex regularisation methods in Banach spaces were discussed and compared in the case of quadratic regularisation on Hilbert spaces. It turned out that, in this simple setting, they were equivalent. In more complicated situations, however, in particular for the regularisation of non-linear operators, but also for

non-convex regularisation (see [11]), variational inequalities might be easier to handle. It is, however, necessary to show that they can also be used for deriving fast convergence rates.

In this paper, it is shown that this is indeed possible for linear operators on Banach spaces. The proof and also the formulation of the variational inequality are based on a duality argument that shows that, if the standard range condition $\text{Ran } A^* \cap \partial\mathcal{R}(x^\dagger)$ holds, then the minimisation of the Tikhonov functional \mathcal{T}_α with noisy data y^δ can be translated to the approximative minimisation of a dual Tikhonov functional with exact data. In the special case of a sufficiently smooth regularisation term \mathcal{R} with sufficiently smooth dual \mathcal{R}^* , the ensuing condition for the best possible rates with this method is almost identical to the range condition for the Hilbert space corresponding to the case $\nu = 1$ in (1).

2 Notation

Let X and Y be Banach spaces and $A: X \rightarrow Y$ a bounded linear operator. Moreover assume that $\mathcal{R}: X \rightarrow [0, +\infty]$ is a convex, lower semi-continuous, and coercive regularisation functional and $p > 1$. Consider, for given data $y \in Y$ and a regularisation parameter $\alpha > 0$, the Tikhonov functional $\mathcal{T}(\cdot; \alpha, y): X \rightarrow [0, +\infty]$ defined by

$$\mathcal{T}_\alpha(x; y) := \frac{1}{p} \|Ax - y\|^p + \alpha\mathcal{R}(x) .$$

From now on we assume that $y^\dagger \in Y$ are some fixed data, and we want to solve the equation

$$Ax = y^\dagger$$

for x . By

$$x^\dagger := \arg \min \{ \mathcal{R}(x) : Ax = y^\dagger \}$$

we denote any \mathcal{R} -minimising solution of the equation $Ax = y^\dagger$. Moreover, for $\alpha > 0$ we denote by

$$x_\alpha := \arg \min \{ \mathcal{T}_\alpha(x; y^\dagger) : x \in X \}$$

any minimiser of the Tikhonov functional with exact data y^\dagger . Finally, for $\delta > 0$ and $\alpha > 0$ we denote by

$$x_\alpha^\delta := \arg \min \{ \mathcal{T}_\alpha(x; y^\delta) : x \in X \} ,$$

any minimiser of the Tikhonov functional with perturbed data y^δ , where $y^\delta \in Y$ is any element satisfying

$$\|y^\delta - y^\dagger\| \leq \delta .$$

The main goal of this paper is the derivation of *convergence rates*, that is, estimates of the distance between x_α^δ and x^\dagger depending on the noise level δ and the regularisation parameter α . Starting with [3], such estimates have been usually derived in terms of the Bregman distance defined by the convex regularisation functional \mathcal{R} .

Definition 2.1. Assume that $x \in X$ and $\xi \in \partial\mathcal{R}(x)$. Then we define the *Bregman distance* with respect to \mathcal{R} as the mapping $\mathcal{D}_\xi(\cdot; x): X \rightarrow [0, +\infty]$,

$$\mathcal{D}_\xi(\tilde{x}; x) = \mathcal{R}(\tilde{x}) - \mathcal{R}(x) - \langle \xi, \tilde{x} - x \rangle.$$

That is, the Bregman distance measures the difference between the graph of \mathcal{R} and its affine approximation at x .

In addition, if $x, \tilde{x} \in X$ and $\xi \in \partial\mathcal{R}(x)$, $\tilde{\xi} \in \partial\mathcal{R}(\tilde{x})$, then we denote by

$$\mathcal{D}_{\xi, \tilde{\xi}}^{\text{sym}}(x, \tilde{x}) := \mathcal{D}_\xi(\tilde{x}; x) + \mathcal{D}_{\tilde{\xi}}(x; \tilde{x})$$

the *symmetrical Bregman distance* between x and \tilde{x} with respect to \mathcal{R} . An easy computation shows that the identity

$$\mathcal{D}_{\xi, \tilde{\xi}}^{\text{sym}}(x, \tilde{x}) = \langle \xi - \tilde{\xi}, x - \tilde{x} \rangle$$

holds. ■

Note that the Bregman distance at x depends on the choice of the subgradient $\xi \in \partial\mathcal{R}(x)$ (unless $\partial\mathcal{R}(x)$ contains a single element). Moreover it can happen that $\partial\mathcal{R}(x) = \emptyset$, in which case the Bregman distance cannot be defined.

3 Duality Mappings

In addition to the Bregman distance with respect to \mathcal{R} , we will also need the Bregman distance with respect to the p -th power of the norm on the space Y . This Bregman distance can be written in terms of duality mappings. In this section we recall several results concerning duality mappings that will be needed for the derivation and interpretation of the main result on convergence rates.

Throughout the whole text, whenever $q > 1$ we denote by q_* its conjugate defined by $1/q + 1/q_* = 1$.

Definition 3.1. Let $q > 1$. The *duality mapping* $J_q: Y \rightarrow 2^{Y^*}$ is defined by

$$J_q(y) = \{ \omega \in Y^* : \langle \omega, y \rangle = \|\omega\| \|y\|, \|\omega\| = \|y\|^{q-1} \}.$$

The *adjoint duality mapping* $J_{q_*}^*: Y^* \rightarrow 2^Y$ is defined by

$$J_{q_*}^*(\omega) = \{ y \in Y : \langle \omega, y \rangle = \|\omega\| \|y\|, \|y\| = \|\omega\|^{q_*-1} \}. \quad \blacksquare$$

If Y is reflexive, one can show that $J_{q_*}^* = (J_q)^{-1}$ (see [5, Chap. 1, Thm. 4.4]). Moreover it is easy to see that J_q is $(q-1)$ -homogeneous. That is, if $y \in Y$ and $\lambda \in \mathbb{R}$, then

$$J_q(\lambda y) = |\lambda|^{q-1} J_q(y).$$

Lemma 3.2. Let $q > 1$ and denote by $\mathcal{S}_q: Y \rightarrow [0, +\infty]$ the mapping

$$\mathcal{S}_q(y) = \frac{1}{q} \|y\|^q.$$

Then

$$J_q(y) = \partial\mathcal{S}_q(y).$$

Proof. See [5, Chap. 2, Cor. 3.5]. □

Recall that the Banach space Y is *smooth* (see [18]), if for every $y \in Y$ satisfying $\|y\| = 1$ there exists a unique $\omega \in Y^*$ such that $\|\omega\| = \langle \omega, y \rangle = 1$. Equivalently, Y is smooth, if and only if every duality mapping J_q on Y with $q > 1$ is single valued. In this case, we regard J_q as a mapping from Y to Y^* . This allows us to formulate the Bregman distance with respect to \mathcal{S}_q in terms of the duality mapping J_q .

Definition 3.3. Assume that Y is a smooth Banach space and let $q > 1$. For $y \in Y$ we define the Bregman distance with respect to \mathcal{S}_q at y by

$$\mathcal{D}_q(\tilde{y}; y) := \frac{1}{q}\|\tilde{y}\|^q - \frac{1}{q}\|y\|^q - \langle J_q(y), \tilde{y} - y \rangle.$$

Moreover we define the symmetric Bregman distance with respect to \mathcal{S}_q as

$$\mathcal{D}_q^{\text{sym}}(\tilde{y}, y) := \langle J_q(\tilde{y}) - J_q(y), \tilde{y} - y \rangle.$$
■

The following definition is taken from [18].

Definition 3.4. The modulus of convexity $\delta_Y : (0, 2] \rightarrow [0, 1]$ is defined by

$$\delta_Y(\varepsilon) := \inf\{1 - \|y - \tilde{y}\|/2 : y, \tilde{y} \in Y, \|y\| = \|\tilde{y}\| = 1, \|y - \tilde{y}\| = \varepsilon\}.$$

The modulus of smoothness $\rho_Y : (0, +\infty) \rightarrow (0, +\infty)$ is defined by

$$\rho_Y(\tau) = \sup\{(\|y + \tilde{y}\| + \|y - \tilde{y}\|)/2 - 1 : y, \tilde{y} \in Y, \|y\| = 1, \|\tilde{y}\| = \tau\}.$$

The space Y is called *uniformly convex* if $\delta_Y(\varepsilon) > 0$ for every $\varepsilon > 0$. It is called *uniformly smooth* if $\rho_Y(\tau) = o(\tau)$ as $\tau \rightarrow 0$.

The space Y is called *q-convex* (or convex of power type q), if there exists a constant $K > 0$ such that $\delta_Y(\varepsilon) \geq K\varepsilon^q$ for all ε . Similarly, it is called *q-smooth* (or smooth of power type q), if $\rho_Y(\tau) \leq K\tau^q$ for all sufficiently small $\tau > 0$. ■

One can show that the space Y is *q-convex*, if and only if Y^* is *q*-smooth* (in fact, the relation $2\rho_{Y^*} = (2\delta_Y)^*$ holds, see [18, Prop. 1.e.2]).

For us, the most important property of *q-convex* spaces is that they allow to estimate the *q*-th power of the norm from above by the Bregman distance. More precisely, the following result holds:

Lemma 3.5. *Assume that the Banach space Y is *q-convex*. Then there exists a constant $C > 0$ such that*

$$C\|y - \tilde{y}\|^q \leq \mathcal{D}_q(\tilde{y}; y).$$

for all $y, \tilde{y} \in Y$.

Proof. See [1, Lemma 2.7]. □

4 Convergence Rates

In the following we always assume that x^\dagger satisfies a range condition with

$$\xi^\dagger = A^* \omega^\dagger \in \text{Ran } A^* \cap \partial \mathcal{R}(x^\dagger).$$

Then the definition of the dual $\mathcal{R}^*: X^* \rightarrow (-\infty, +\infty]$ of \mathcal{R} implies that

$$x^\dagger \in \partial \mathcal{R}^*(A^* \omega^\dagger).$$

Consequently we can define a Bregman distance $\mathcal{D}_{x^\dagger}^*(\cdot; A^* \omega^\dagger): X^* \rightarrow [0, +\infty]$ for the dual function \mathcal{R}^* as

$$\mathcal{D}_{x^\dagger}^*(\xi; A^* \omega^\dagger) = \mathcal{R}^*(\xi) - \mathcal{R}^*(A^* \omega^\dagger) - \langle \xi - A^* \omega^\dagger, x^\dagger \rangle.$$

Proposition 4.1. *Assume that x_α^δ minimises $\mathcal{T}_\alpha(\cdot; y^\delta)$. Then there exists $\omega_\alpha^\delta \in Y^*$ such that*

$$A^* \omega_\alpha^\delta \in \partial \mathcal{R}(x_\alpha^\delta), \quad -\alpha \omega_\alpha^\delta = J_p(Ax_\alpha^\delta - y^\delta). \quad (2)$$

Moreover ω_α^δ minimises the functional $\mathcal{T}_\alpha^*(\cdot; y^\delta): Y^* \rightarrow (-\infty, +\infty]$ defined by

$$\mathcal{T}_\alpha^*(\omega; y^\delta) := \mathcal{D}_{x^\dagger}^*(A^* \omega; A^* \omega^\dagger) + \alpha^{p_*-1} \frac{1}{p_*} \|\omega\|^{p_*} - \langle \omega - \omega^\dagger, y^\delta - y^\dagger \rangle.$$

Proof. The dual of the problem

$$\mathcal{T}_\alpha(x)/\alpha = \frac{1}{\alpha p} \|Ax - y^\delta\|^p + \mathcal{R}(x) \rightarrow \min$$

is the problem

$$\frac{1}{\alpha p_*} \|\alpha \omega\|^{p_*} - \langle \omega, y^\delta \rangle + \mathcal{R}^*(A^* \omega) \rightarrow \min$$

(see [6, Chap. III]). Writing

$$-\langle \omega, y^\delta \rangle = \langle \omega, y^\dagger - y^\delta \rangle - \langle \omega, y^\dagger \rangle = -\langle \omega, y^\delta - y^\dagger \rangle - \langle A^* \omega, x^\dagger \rangle$$

and adding the constant terms $\langle A^* \omega^\dagger, x^\dagger \rangle$, $\langle \omega^\dagger, y^\delta - y^\dagger \rangle$, and $-\mathcal{R}^*(A^* \omega^\dagger)$, we obtain the problem

$$\begin{aligned} \alpha^{p_*-1} \frac{1}{p_*} \|\omega\|^{p_*} + \mathcal{R}^*(A^* \omega) - \mathcal{R}^*(A^* \omega^\dagger) \\ - \langle A^* \omega - A^* \omega^\dagger, x^\dagger \rangle - \langle \omega - \omega^\dagger, y^\delta - y^\dagger \rangle \rightarrow \min, \end{aligned}$$

which is precisely the minimisation of the dual Tikhonov functional \mathcal{T}_α^* . The relations (2) are nothing else than the Karush–Kuhn–Tucker conditions for the minimisation problem (cf. [6, Chap. III]). \square

Remark 4.2. In particular, Proposition 4.1 implies that, if x_α^δ minimises the primal Tikhonov functional with noisy data y^δ , then ω_α^δ *almost* minimises the dual Tikhonov functional $\mathcal{T}_\alpha^*(\cdot; y^\dagger)$ with *exact* data y^\dagger . More precisely, we have

$$\mathcal{T}_\alpha^*(\omega_\alpha^\delta; y^\dagger) \leq \inf \{ \mathcal{T}_\alpha^*(\omega; y^\dagger) : \omega \in Y^* \} + \delta \|\omega_\alpha^\delta - \omega^\dagger\|. \quad \blacksquare$$

Definition 4.3. An *index function* is a strictly increasing, continuous and concave function $\Phi: [0, +\infty) \rightarrow [0, +\infty)$ satisfying $\Phi(0) = 0$. \blacksquare

Theorem 4.4. Assume that Y is a p -smooth Banach space and that there exists an index function Φ such that

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq \Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega; A^*\omega^\dagger)) \quad (3)$$

for every $\omega \in Y^*$. Denote by Ψ the conjugate of the convex mapping $t \mapsto \Phi^{-1}(t)$. Then there exists a constant D only depending on p and the Banach space Y such that

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) \leq \Psi(\alpha^{p_*-1}) + D \frac{\delta^p}{\alpha}.$$

Proof. We have

$$\begin{aligned} \mathcal{D}_{A^*\omega_\alpha^\delta, A^*\omega^\dagger}^*(x^\dagger, x_\alpha^\delta) &= \langle A^*\omega_\alpha^\delta - A^*\omega^\dagger, x_\alpha^\delta - x^\dagger \rangle \\ &= \langle \omega_\alpha^\delta - \omega^\dagger, Ax_\alpha^\delta - Ax^\dagger \rangle \\ &= \langle \omega_\alpha^\delta - \omega^\dagger, Ax_\alpha^\delta - y^\delta \rangle + \langle \omega_\alpha^\delta - \omega^\dagger, y^\delta - y^\dagger \rangle. \end{aligned}$$

Proposition 4.1 implies that $-\alpha\omega_\alpha^\delta = J_p(Ax_\alpha^\delta - y^\delta)$, and therefore

$$Ax_\alpha^\delta - y^\delta = J_{p_*}(-\alpha\omega_\alpha^\delta) = -\alpha^{p_*-1}J_{p_*}(\omega_\alpha^\delta).$$

Consequently

$$\begin{aligned} \mathcal{D}_{A^*\omega_\alpha^\delta, A^*\omega^\dagger}^*(x^\dagger, x_\alpha^\delta) &= -\alpha^{p_*-1}\langle \omega_\alpha^\delta - \omega^\dagger, J_{p_*}(\omega_\alpha^\delta) \rangle + \langle \omega_\alpha^\delta - \omega^\dagger, y^\delta - y^\dagger \rangle \\ &= -\alpha^{p_*-1}\langle \omega_\alpha^\delta - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle - \alpha^{p_*-1}\mathcal{D}_{p_*}^*(\omega_\alpha^\delta, \omega^\dagger) \\ &\quad + \langle \omega_\alpha^\delta - \omega^\dagger, y^\delta - y^\dagger \rangle \end{aligned}$$

Using (3) and estimating $\langle \omega_\alpha^\delta - \omega^\dagger, y^\delta - y^\dagger \rangle \leq \delta \|\omega_\alpha^\delta - \omega^\dagger\|$, we obtain

$$\begin{aligned} \mathcal{D}_{A^*\omega_\alpha^\delta, A^*\omega^\dagger}^*(x^\dagger, x_\alpha^\delta) &\leq \alpha^{p_*-1}\Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger)) - \alpha^{p_*-1}\mathcal{D}_{p_*}^*(\omega_\alpha^\delta, \omega^\dagger) + \delta \|\omega_\alpha^\delta - \omega^\dagger\|. \quad (4) \end{aligned}$$

Now,

$$\mathcal{D}_{A^*\omega_\alpha^\delta, A^*\omega^\dagger}^*(x^\dagger, x_\alpha^\delta) = \mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) + \mathcal{D}_{A^*\omega_\alpha^\delta}(x^\dagger; x_\alpha^\delta),$$

and the assumptions $A^*\omega^\dagger \in \partial\mathcal{R}(x^\dagger)$ and $A^*\omega_\alpha^\delta \in \partial\mathcal{R}(x_\alpha^\delta)$ imply that

$$\begin{aligned} \mathcal{D}_{A^*\omega_\alpha^\delta}(x^\dagger; x_\alpha^\delta) &= \mathcal{R}(x^\dagger) - \mathcal{R}(x_\alpha^\delta) - \langle A^*\omega_\alpha^\delta, x^\dagger - x_\alpha^\delta \rangle \\ &= \langle A^*\omega^\dagger, x^\dagger \rangle - \mathcal{R}^*(A^*\omega^\dagger) - \langle A^*\omega_\alpha^\delta, x_\alpha^\delta \rangle + \mathcal{R}^*(A^*\omega_\alpha^\delta) \\ &\quad - \langle A^*\omega_\alpha^\delta, x^\dagger - x_\alpha^\delta \rangle \\ &= \mathcal{R}^*(A^*\omega_\alpha^\delta) - \mathcal{R}^*(\omega^\dagger) - \langle A^*(\omega_\alpha^\delta - \omega^\dagger), x^\dagger \rangle \\ &= \mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger). \end{aligned}$$

Because Y is p -smooth, its dual space Y^* is p_* -convex. Thus there exists $C > 0$ such that (see Lemma 3.5)

$$C \|\omega_\alpha^\delta - \omega^\dagger\|^{p_*} \leq \mathcal{D}_{p_*}^*(\omega_\alpha^\delta, \omega^\dagger).$$

Consequently (4) implies that

$$\begin{aligned}\mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) &\leq \alpha^{p_*-1} \Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger)) - \mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger) \\ &\quad + \delta \|\omega_\alpha^\delta - \omega^\dagger\| - C\alpha^{p_*-1} \|\omega_\alpha^\delta - \omega^\dagger\|^{p_*}.\end{aligned}$$

Applying Young's inequality (see [15, Thm. 13.2]), we obtain

$$\begin{aligned}\alpha^{p_*-1} \Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger)) &\leq \Psi(\alpha^{p_*-1}) + \Psi^*(\Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger))) \\ &= \Psi(\alpha^{p_*-1}) + \mathcal{D}_{x^\dagger}^*(A^*\omega_\alpha^\delta; A^*\omega^\dagger)\end{aligned}$$

and

$$\delta \|\omega_\alpha^\delta - \omega^\dagger\| \leq C\alpha^{p_*-1} \|\omega_\alpha^\delta - \omega^\dagger\|^{p_*} + \frac{1}{pp_*^{1/p_*} C^{p/p_*}} \frac{\delta^p}{\alpha}.$$

Therefore

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) \leq \Psi(\alpha^{p_*-1}) + \frac{1}{pp_*^{1/p_*} C^{p/p_*}} \frac{\delta^p}{\alpha},$$

which proves the assertion. \square

Corollary 4.5. *Assume that the assumptions of Theorem 4.4 are satisfied. Then we have for exact data y^\dagger the estimate*

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha; x^\dagger) \leq \Psi(\alpha^{p_*-1}).$$

In the special case where $\Phi(t) = ct^{1/q_}$ for some $q_* > 1$ and $c > 0$ we have*

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha; x^\dagger) \leq \frac{c^q q_*^{1-q}}{q} \alpha^{(p_*-1)q}.$$

Proof. In the general case, the assertion directly follows from Theorem 4.4 for $\delta = 0$. In the case $\Phi(t) = ct^{1/q_*}$, the mapping Ψ is the conjugate of $\Phi^{-1}(t) = t^{q_*}/c^{q_*}$. A short calculation shows that Ψ indeed has the form stated above. \square

Corollary 4.6. *Assume that the assumptions of Theorem 4.4 are satisfied with an index function $\Phi(t) \sim t^{1/q_*}$ for some $q_* > 1$. Then we have for a parameter choice*

$$\alpha \sim \delta^{\frac{p}{(p_*-1)q+1}}$$

the convergence rate

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) = O\left(\delta^{\frac{p_*q}{(p_*-1)q+1}}\right).$$

Proof. From Theorem 4.4 we obtain the estimate

$$\mathcal{D}_{A^*\omega^\dagger}(x_\alpha^\delta; x^\dagger) \leq \Psi(\alpha^{p_*-1}) + D \frac{\delta^p}{\alpha},$$

where Ψ is the conjugate of the function Φ^{-1} . Since $\Phi(t) \sim t^{1/q_*}$, it follows that $\Phi^{-1} \sim t^{q_*}$, and therefore $\Psi \sim t^q$. Thus

$$\Psi(\alpha^{p_*-1}) + D \frac{\delta^p}{\alpha} \sim \alpha^{(p_*-1)q} + \frac{\delta^p}{\alpha} \sim \delta^{\frac{p(p_*-1)q}{(p_*-1)q+1}} + \delta^{p-\frac{p}{(p_*-1)q+1}} \sim \delta^{\frac{p(p_*-1)q}{(p_*-1)q+1}},$$

which proves the assertion, as $p(p_* - 1) = p_*$. \square

5 Implications of Smoothness

In this section, we first discuss the limitations of Theorem 4.4 in the case where the dual of \mathcal{R} is smooth. Then the best possible rates imply that the dual variable ω^\dagger satisfies the range condition $J_{p_*}(\omega^\dagger) \in \text{Ran } A$. Conversely, this range condition implies rates, if the functional \mathcal{R} itself is sufficiently smooth.

Lemma 5.1. *Assume that \mathcal{R}^* is two times Fréchet differentiable at $A^*\omega^\dagger$ and that there exists an index function Φ such that*

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq \Phi(\mathcal{D}_{x^\dagger}^*(A^*\omega; A^*\omega^\dagger)) \quad (5)$$

for all $\omega \in Y^*$. If $\Phi(t) = o(t^{1/2})$ as $t \rightarrow 0$, then x^\dagger minimises \mathcal{R} .

In addition, if one can choose $\Phi(t) \sim t^{1/2}$ as $t \rightarrow 0$, then

$$J_{p_*}(\omega^\dagger) \in \text{Ran } A .$$

Proof. Because \mathcal{R}^* is two times Fréchet differentiable at $A^*\omega^\dagger$, a Taylor expansion of \mathcal{R}^* around $A^*\omega^\dagger$ shows that

$$\begin{aligned} \mathcal{D}_{x^\dagger}^*(A^*\omega; A^*\omega^\dagger) &= \mathcal{R}^*(A^*\omega) - \mathcal{R}^*(A^*\omega^\dagger) - \langle x^\dagger, A^*(\omega - \omega^\dagger) \rangle \\ &= (\mathcal{R}^*)''(A^*\omega^\dagger)(A^*(\omega - \omega^\dagger); A^*(\omega - \omega^\dagger)) + o(\|A^*(\omega - \omega^\dagger)\|^2), \end{aligned}$$

where $(\mathcal{R}^*)''(A^*\omega^\dagger) \in B(X^*)$ denotes the second order derivative of \mathcal{R}^* at $A^*\omega^\dagger$, which is a symmetric, bounded bilinear form on X^* . Writing $\tilde{\omega} := \omega - \omega^\dagger$, the inequality (5) implies that

$$\langle \tilde{\omega}, J_{p_*}(\omega^\dagger) \rangle \leq \Phi((\mathcal{R}^*)''(A^*\omega^\dagger)(A^*\tilde{\omega}; A^*\tilde{\omega}) + o(\|A^*\tilde{\omega}\|^2))$$

for all $\tilde{\omega} \in X^*$. In addition, one can estimate

$$(\mathcal{R}^*)''(A^*\tilde{\omega}; A^*\tilde{\omega}) \leq \|(\mathcal{R}^*)''(A^*\omega^\dagger)\| \|A^*\tilde{\omega}\|^2 .$$

Therefore,

$$\langle \tilde{\omega}, J_{p_*}(\omega^\dagger) \rangle \leq \Phi(\|(\mathcal{R}^*)''(A^*\omega^\dagger)\| \|A^*\tilde{\omega}\|^2 + o(\|A^*\tilde{\omega}\|^2)) . \quad (6)$$

Now assume that $\Phi(t) = o(t^{1/2})$ as $t \rightarrow 0$. Then, dividing (6) by $\|A^*\tilde{\omega}\|$ and considering the limit $\|A^*\tilde{\omega}\| \rightarrow 0$, we see that $J_{p_*}(\omega^\dagger) = 0$, which is equivalent to stating that $\omega^\dagger = 0$. Thus $0 = A^*\omega^\dagger \in \partial\mathcal{R}(x^\dagger)$, which proves that x^\dagger minimises \mathcal{R} .

Now assume that $\Phi(t) \sim t^{1/2}$ as $t \rightarrow 0$. Then (6) implies that there exists a constant $C > 0$ such that

$$\langle \tilde{\omega}, J_{p_*}(\omega^\dagger) \rangle \leq C(\|(\mathcal{R}^*)''(A^*\omega^\dagger)\|)^{1/2} \|A^*\tilde{\omega}\|$$

for $\|A^*\tilde{\omega}\|$ sufficiently small—and thus everywhere, as the right hand side is positively homogeneous and the left hand side is linear. Thus [24, Lemma 8.21] implies that $J_{p_*}(\omega^\dagger) \in \text{Ran } A$. \square

In the special case, where X is a q -convex Banach space and

$$\mathcal{R}(x) = \frac{1}{q} \|x\|^q ,$$

the above result can be slightly refined. In this case, the condition of Theorem 4.4 reads as

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq \Phi(\mathcal{D}_{q_*}(A^* \omega; A^* \omega^\dagger))$$

and the resulting estimate is

$$\mathcal{D}_q(x_\alpha^\delta; x^\dagger) \leq \Psi(\alpha^{p_*-1}) + D \frac{\delta^p}{\alpha}.$$

Lemma 5.2. *Assume that X is q -convex and*

$$\mathcal{R}(x) = \frac{1}{q} \|x\|^q,$$

and that

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq \Phi(\mathcal{D}_{q_*}(A^* \omega; A^* \omega^\dagger))$$

for some index function Φ . If

$$\Phi(t) = o(t^{1/q_*}) \quad \text{as } q \rightarrow 0,$$

then $x^\dagger = 0$.

Proof. We have

$$\Phi(\mathcal{D}_{q_*}(A^* \omega; A^* \omega^\dagger)) \leq \Phi(\hat{C}_{q_*} \|A^*(\omega - \omega^\dagger)\|^{q_*}).$$

Thus the assumption $\Phi(t) = o(t^{1/q_*})$ implies that, writing $\tilde{\omega} := \omega - \omega^\dagger$,

$$\frac{\langle \tilde{\omega}, J_{p_*}(\omega^\dagger) \rangle}{\|\tilde{\omega}\|} \leq \frac{\Phi(\hat{C}_{q_*} \|A^*\tilde{\omega}\|^{q_*})}{\|\tilde{\omega}\|} \rightarrow 0 \quad \text{as } \tilde{\omega} \rightarrow 0.$$

This, however, is only possible, if $J_{p_*}(\omega^\dagger) = 0$, which is equivalent to stating that $\omega^\dagger = 0$. Thus $0 = A^*\omega^\dagger \in J_q(x^\dagger)$, and therefore $x^\dagger = 0$. \square

Lemma 5.3. *Assume that \mathcal{R} is two times Fréchet differentiable at x^\dagger and that $J_{p_*}(\omega^\dagger) \in \text{Ran } A$. Then there exists $C > 0$ such that*

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq C(\mathcal{D}_{x^\dagger}^*(A^* \omega; A^* \omega^\dagger))^{1/2}$$

for ω sufficiently close to ω^\dagger .

Proof. Because \mathcal{R} is two times Fréchet differentiable at x^\dagger , there exists $c > 0$ such that

$$\mathcal{R}(x) \leq \mathcal{R}(x^\dagger) + \langle A^*\omega^\dagger, x - x^\dagger \rangle + \frac{c}{2} \|x - x^\dagger\|^2$$

for every $x \in X$ sufficiently close to x^\dagger . Consequently,

$$\mathcal{R}^*(A^* \omega) \geq \mathcal{R}^*(A^* \omega^\dagger) + \langle A^*(\omega - \omega^\dagger), x^\dagger \rangle + \frac{1}{2c} \|A^*(\omega - \omega^\dagger)\|^2$$

for ω sufficiently close to ω^\dagger . This, however, is equivalent to stating that, locally around ω^\dagger ,

$$\mathcal{D}_{x^\dagger}^*(A^* \omega; A^* \omega^\dagger) \geq \frac{1}{2c} \|A^*(\omega - \omega^\dagger)\|^2.$$

Now the assumption that $J_{p_*}(\omega^\dagger) \in \text{Ran } A$ is equivalent to the estimate

$$\langle \omega - \omega^\dagger, J_{p_*}(\omega^\dagger) \rangle \leq \tilde{c} \|A^*(\omega - \omega^\dagger)\|$$

for some $\tilde{c} > 0$ and all $\omega \in Y^*$. Assembling these inequalities, the assertion follows. \square

6 Conclusion

In this paper it is shown that the approach of variational inequalities can be used for the derivation of higher order convergence rates and is thus not restricted to the “low rate world” as has been surmised in [10]. The basic idea for this generalisation is the formulation of the variational inequality not for the primal Tikhonov functional, but rather for a dual functional. By this approach we obtain the whole range of convergence rates that have already been derived for quadratic regularisation and also for convex regularisation on Banach spaces. The main advantage of the usage of variational inequalities is their comparative simplicity, in particular when used in conjunction with non-linear operators; they have been introduced precisely for the study of non-linear ill-posed operator equations. Thus it seems reasonable that the approach can be extended also to the non-linear case without introducing too many artificial constraints on the operator.

References

- [1] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functionals in Banach spaces. *Abstr. Appl. Anal.*, 2008:Art. ID 192679, 19, 2008.
- [2] R.I. Boț and B. Hofmann. An extension of the variational approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *J. Integral Equations Appl.*, 22:369–392, 2010.
- [3] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.
- [4] G. Chavent and K. Kunisch. Convergence of Tikhonov regularization for constrained ill-posed inverse problems. *Inverse Probl.*, 10:63–76, 1994.
- [5] I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, volume 62 of *Mathematics and its Applications*. Kluwer, Dordrecht, 1990.
- [6] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels*. Dunod, 1974. Collection Études Mathématiques.
- [7] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [8] H. W. Engl and G. Landl. Convergence rates for maximum entropy regularization. *SIAM J. Numer. Anal.*, 30(5):1509–1536, 1993.
- [9] J. Flemming. Solution smoothness of ill-posed equations in Hilbert spaces: four concepts and their cross connections. *Appl. Anal.*, iFirst:1–16, 2011. DOI: 10.1080/00036811.2011.563736.
- [10] J. Flemming and B. Hofmann. A new approach to source conditions in regularization with general residual term. *Numer. Funct. Anal. Optim.*, 31(3):245–284, 2010.
- [11] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Probl.*, 26(11):115014, 2010.
- [12] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, Boston, 1984.

- [13] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Probl.*, 24(4):045007, 10, 2008.
- [14] T. Hein. Tikhonov regularization in Banach spaces: improved convergence rates results. *Inverse Problems*, 25(3):035002 (18pp), 2009.
- [15] E. Hewitt and K. Stromberg. *Real and Abstract Analysis*. Springer Verlag, New York, 1965.
- [16] B. Hofmann. Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators. *Math. Methods Appl. Sci.*, 29(3):351–371, 2006.
- [17] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Probl.*, 23(3):987–1010, 2007.
- [18] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1979. Function spaces.
- [19] A. Neubauer. Finite-dimensional approximation of constrained Tikhonov-regularized solutions of ill-posed linear operator equations. *Math. Comp.*, 48(178):565–583, 1987.
- [20] A. Neubauer. On enhanced convergence rates for Tikhonov regularization of non-linear ill-posed problems in Banach spaces. *Inverse Probl.*, 25(6):065009 (10pp), 2009.
- [21] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Appl. Anal.*, 89(11):1729–1743, 2010.
- [22] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Probl.*, 21(4):1303–1314, 2005.
- [23] O. Scherzer. A posteriori error estimates for the solution of nonlinear ill-posed operator equations. *Nonlinear Anal.*, 45(4, Ser. A: Theory Methods):459–481, 2001.
- [24] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.